UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

Nonparametric Bayesian Models for Signal Processing

by

Caoyuan Li

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Sydney, Australia

2019

Certificate of Authorship/Originality

I certify that this thesis has been written by me. Any help that I have received in my research and in the preparation of the thesis itself has been fully acknowledged. In addition, I certify that all information sources and literature used are quoted in the thesis.

This thesis is the result of the research candidature conducted jointly with Beijing Institute of Technology as part of a collaborative doctoral degree.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Signature of Student: ____________

Date: 23/06/2020

ABSTRACT

Nonparametric Bayesian Models for Signal Processing

by

Caoyuan Li

An essential component in signal processing is to remove various kinds of noise from the signal. It is possible to introduce noise during the process of signal storage, transmission and acquisition. Signal quality after denoising affects subsequent signal analysis profoundly. Low-rank representation is a popular method in signal processing. It is aimed to capture underlying low-dimensional structures of high dimensional signal and attracted much attention in the area of the pattern recognition and signal processing. Such successful applications were mainly due to its effectiveness in exploring low dimensional manifolds embedded in data, which can be naturally characterized by low rankness of the data matrix.

This thesis conducts research on processing various signals as well as getting the low-rank representation of the signal via the variational Bayesian inference techniques. This study proposed four different nonparametric Bayesian models for image denoising, inpainting, video foreground/background separation and bio-medical signal processing as follows.

(1) A hybrid denoising model based on variational Bayesian inference and Stein's unbiased risk estimator (SURE) is presented, which consists of two complementary steps. In the first step, the variational Bayesian singular value thresholding (SVT) performs a low-rank approximation of the nonlocal image patch matrix to simultaneously remove the noise and estimate the noise variance. In the second step, the conventional SURE full rank SVT and its divergence formulas for rank-reduced eigen-triplets is modified to remove the residual artefacts.

(2) A hierarchical kernelized sparse Bayesian matrix factorization (KSBMF)

model is developed to integrate side information. The KSBMF automatically infers the parameters and latent variables including the reduced rank using the variational Bayesian inference. Also, the model simultaneously achieves low-rankness through sparse Bayesian learning and sparsity through an enforced constraint on latent factor matrices. The KSBMF is further connected with the nonlocal image processing framework to develop two algorithms for image denoising and inpainting.

(3) A robust kernelized Bayesian matrix factorization (RKBMF) model is proposed to decompose a data set into low rank and sparse components. Moreover, the model integrates the side information of similarity between frames to improve information extraction from the video. RKBMF is employed to extract background and foreground information from a traffic video.

(4) A hierarchical Dirichlet process nonnegative matrix factorization (DPNMF) model is presented in which the Gaussian mixture model is used to approximate the complex noise distribution. Moreover, the model is cast in the nonparametric Bayesian framework by using Dirichlet process mixture to infer the necessary number of Gaussian components. A mean-field variational inference algorithm is derived for the proposed nonparametric Bayesian model. The model is tested on synthetic data sets contaminated by Gaussian, sparse and mixed noise. The proposed model is then applied to extract muscle synergies from the electromyographic (EMG) signal and to select discriminative features for motor imagery single-trial electroencephalogram (EEG) classification.

Dissertation directed by Associate Professor Richard Xu
School of Electrical and Data Engineering

Dedication

To my parents and my wife for your love and support.

Acknowledgements

The completion of this dissertation has been possible with the inspiration and encouragement from many people, to whom I am greatly indebted.

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Richard Yi Da Xu, for the continuous support of my PhD study and research in UTS. He is my most admired supervisor. He treats each student as his best friend, trying his best to benefit us and always thinking from the student's perspective. I will never forget the scene where we drink beer together at the bar and the nights we spend together to improve my papers. I gained a lot of knowledge from him, not only the knowledge of machine learning but also experience about how to deal with people and how to survive in the workplace. I believe that I will continue to benefit from this experience for the rest of my life. His guidance helped me in all the time of research and writing of this thesis. I would thank him, for sharing his immense knowledge with me, keep encouraging and motivating me. Without his professional guidance and persistent help, this thesis would not have been possible.

I also would like to appreciate my co-supervisor Dr Xuhui Fan and Dr Hong-bo Xie, for sharing me research ideas and their invaluable experience about research. I wouldn't have finished this thesis without their selfless help.

I thank my fellow labmates in UTS: Xuan Liang, Shuai Jiang, Haodong Chang, Wanming Huang, Wei Huang, etc., for our stimulating discussions and for all the fun we have had in the last several years.

Also, I would like to thank the magic of machine learning. This fantastic world has made me lucky enough to have fun while exploring it during all of my PhD period.

Last but not least, I would like to thank my family: my wife, my parents and my parents in law, for their unconditional support, both financially and emotionally throughout the whole PhD studying.

Caoyuan Li

Sydney, Australia, 2019.

List of Publications

Journal Papers

J-1. **Caoyuan Li**, Hongbo Xie, Xuhui Fan, Richard Yi Da Xu, et al. Image denoising based on nonlocal Bayesian singular value thresholding and Steins unbiased risk estimator. *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 4899-4911, Oct. 2019

J-2. **Caoyuan Li**, Hongbo Xie, Xuhui Fan, Richard Yi Da Xu, et al. Kernelized sparse Bayesian matrix factorization. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

J-3. **Caoyuan Li**, Hongbo Xie, Kerrie Mengersen, et al. Bayesian nonnegative matrix factorization with Dirichlet process mixtures. *IEEE Transactions on Signal Processing*, 2020.

Conference Papers

C-1. Hongbo Xie, **Caoyuan Li**, Richard Yi Da Xu, et al. Robust kernelized Bayesian matrix factorization for video background/foreground separation. *The Fifth International Conference on Machine Learning, Optimization, and Data Science*, September 10-13, 2019

Contents

List of Figures

List of Tables

Nomenclature and Notation

Example	**Description**
$\mathbb{R}$	The set of reals
$\mathbf{N}$	the set of natural numbers
$\mathbb{E}[\cdot]$	expection of a random variable
$\langle\cdot\rangle$	expection of a random variable
$\boldsymbol{Y} \in \mathbb{R}^{m\times n}$	Bold and capitalized letters denote random matrices
$\mathbf{u}$	Characters in bold denotes random vectors
	This notation is also used to denote collections of random variables
$\mathbf{a}_{m\cdot}$	m-th row of a matrix
$\mathbf{a}_{\cdot n}$	n-th column of a matrix
y	Characters in italic denote random scalars
$diag(\mathbf{x})$	Diagonal matrix with the values of vector $\mathbf{x}$ on the diagonal.
I_n	The identity matrix of dimension $n \times n$
θ	Parameters of a model are typcally denoted with the Greek lowercase letter θ
$(.)^\top$	denotes the transpose operation
$\mathrm{p}(\mathbf{x})$	Probability density functions (PDFs) and probability mass functions (PMFs)
$\mathrm{p}(\mathbf{x}, \mathbf{y}, \mathbf{z})$	Joint distributions are denoted by $p(\cdot,\cdot)$
$p(\mathbf{x}\vert\mathbf{z})$	Condition distributions are denoted by $p(\cdot\vert\cdot)$